Simulation Theory

Shannon Spaulding

(This is a penultimate draft of a paper that will appear in Handbook of Imagination, Amy Kind (ed.). Routledge Press. Please cite only the final printed version.)

1. Introduction

The Simulation Theory is an account of how we understand other people. In brief, the theory holds that we understand others by mentally simulating being them. Observing your behavior, I imagine what I would think, feel, and do in your situation. On the basis of this imaginative simulation, I make an inference about what you think, feel, and will do.

Originally, the Simulation Theory was advanced as a theory of mindreading, i.e., as a theory about how we attribute mental states to others in order to explain and predict their behavior (Goldman, 2006b; Gordon, 1986). The theory has been co-opted to explain pretense (Currie & Ravenscroft, 2002), engagement with fiction (Currie, 2010; Goldman, 2006a), imitation (Hurley, 2005), and other cognitive feats (Davies & Stone, 1995b). In each of these roles, simulation crucially involves imagination. In this chapter, I will focus on the Simulation Theory as a theory of mindreading. However, when appropriate, I will draw connections with these other uses of the theory.

Mindreading consists in attributing a mental state to a target in order to understand the target's behavior and anticipate future behavior. Mindreading theorists typically fall into two distinct camps: the Theory Theory (TT) and the Simulation Theory (ST). Theory theorists argue that we understand others by employing a folk psychological theory to explain and predict others' behaviors (Carruthers & Smith, 1996; Davies & Stone, 1995a). In contrast, simulation theorists argue that we do not need to employ a *theory* about folk psychology to understand others. The theory theorists over-intellectualize this process. To understand a target's behavior, all we need to do is imagine what *we* would think, feel, and do in the target's situation, and on that basis we come to understand what the target thinks, feels, and will do (Davies & Stone, 1995b).

There are two kinds of simulational mindreading, which often work together. The first kind is *retrodictive* simulation. We simulate the target to figure out what the target's mental states could have been to cause the observed behavior. That is, we imagine ourselves in the target's situation and imagine what mental states could have caused us to act in the way the target acts. The result of the retrodictive simulation is an explanation of the target's behavior. The second kind is *projective* simulation, which can take as input the output of the retrodictive simulation. We imagine that we have the target's beliefs and desires, and we imagine what we would do next in that situation. We take the resulting conclusion and attribute it to the target. The result of the projective simulation is a prediction of what the target will do next.

The ST sometimes is characterized as an *information-poor* mindreading process. It does not require access to large bodies of information about folk psychology. We do not need theoretical knowledge about causal relations between mental states and behavior. Simulation requires the ability to imagine oneself in a target's situation and figure out what one would feel, think, and do in that position. One simply redeploys one's own cognitive mechanisms for the purpose of mindreading. It is in this sense that the ST is an information-*poor* mindreading process (Goldman, 2006b).

Simulational mindreading can be either "high-level" or "low-level." For high-level mindreading, the mental states involved are beliefs and other propositional attitudes, and the mindreading process is consciously and voluntarily mediated. For low-level mindreading, the mental states involved are non-propositional attitudes, such as emotions and basic intentions, and the mindreading process is largely automatic and non-conscious.[1]

In this chapter, I discuss three aspects of the ST: the concept of simulation, high-level simulation, and low-level simulation. In the next section, I argue for a more precise characterization of the concept of simulation. In sections 3 and 4, I discuss high-level simulation and low-level simulations, which are quite different in some respects. High-level simulation involves imagination in the conventional sense. The process and product of high-level simulation are consciously accessible and potentially under the control of the agent. In contrast, low-level simulation involves an unconventional sense of imagination, and the process and product of

[1] The distinction between low-level and high-level simulation need not take a stand on propositionalism, the view that all attitudes are propositional. If your view is that that all attitudes are propositional attitudes, then on your view low-level simulation involves propositional attitudes. Throughout the paper I will assume that there are non-propositional attitudes, but my arguments do not hang on this assumption.

low-level simulation are involuntary and not consciously accessible. I discuss the features and potential limitations of both high-level and low-level simulation.

2. The Concept of Simulation

One difficulty in assessing the ST is that the concept of simulation often is not well defined. There is no unified conception of simulation across the sciences. Computer simulations, physical simulations, abstract simulations in scientific modeling, and mental simulations have little in common. There is no consensus on what exactly simulation is even in the cognitive sciences, not even in the area of mindreading. As a result, some theorists have called for a moratorium on the term (Nichols & Stich, 2003, p. 134).

Within the field of mindreading, Alvin Goldman has done the most to advance the concept of simulation. My exploration of the ST will be based on Goldman's influential work. Goldman defines simulation in the following way: Process P simulates process P´ iff (1) P duplicates, replicates or resembles P´ in some significant respect and (2) in its duplication of P´, P fulfills one of its purposes or functions. For mindreading simulation, P and P´ are mental processes and the purpose or function of P – the simulating process – is to understand a target's mental states (Goldman, 2006b).

This definition of simulation is vague.[2] With respect to (1), it is unclear what is it for a process to duplicate, replicate or resemble another process *in some significant respect*. Following Justin Fisher (2006), let us distinguish *concrete* and *abstract* replications. When event A concretely replicates event B, A and B involve the same types of systems and exhibit similar fine-grained details. Building a functional model rocket is an example of concrete replication. The model rocket reproduces the parts of the rocket and the relations among those parts and the environment. An accurate model rocket will replicate, in proportion, the concrete details of the actual rocket. When using a model rocket to explain or predict the behavior of an actual rocket, the success of the replication *depends on* reproducing the fine-grained details of the actual rocket and its environment.

In contrast, when event A abstractly replicates event B, A and B need not involve the same types of systems or fine-grained details. For example, the mathematical process in my head when I add 2 and 3 abstractly resembles a calculator's mathematical process of adding 2 and 3. Both the system underlying the calculator's mathematical process and the fine-grained details of the process differ from the cognitive system underlying, and the fine-grained details of, my mathematical process.

The central difference between abstract and concrete replication is that successful abstract replication does not require sameness of system and fine-

[2] This discussion of the concept of simulation is based on previous work. See Spaulding (2012) for a more in depth discussion.

grained process similarity, whereas successful concrete replication does.[3] The calculator may successfully replicate my mathematical process – that is, it may perform a mathematical process the output of which is the same as the output of the mathematical process that occurs in me – without matching the fine-grained details of my cognitive process. Crucially, when A abstractly replicates B, A and B may in fact be the same type of process and share the same fine-grained details, but the success of the replication – equifinality – does not depend on this similarity. This point will be important later.

I suggest that when Goldman claims that simulation requires that process P replicates, duplicates or resembles process P´ *in some significant respect*, this ought to mean that P *concretely* replicates P´. Abstract replications do not count as significant replications because they are ubiquitous. Using a calculator, an abacus, a look-up table, counting on one's fingers, and the operations of a particular Turing machine are all abstract replications of the mathematical process that occurs in my mind when I add 2 and 3. At a coarse-grained level they resemble the mathematical process that occurs in my brain when I add 2 and 3. In contrast, concrete replications generally are far less common than abstract replications (because the criteria for concrete replications are more stringent), and the replicational aspect plays a central role in their success. If the ST is to offer an interesting, distinctive claim about how mindreading operates, it ought to be restricted to concrete replications.

---

[3] What counts as a successful replication depends on the particular replication at hand. For the mathematical process, it is coming up with the same result. For the rocket, it is duplicating the conditions of the actual rocket.

Another reason that simulation should require concrete replication is that some versions of the TT entail abstract replication (Fisher, 2006, p. 422). For example, in predicting whether you will decide to go to the party on Friday night, I call on my knowledge that you enjoy parties, you like the people who are attending the party, you have had a difficult week and desire to go out, and you know that you have been invited to a party on Friday night. I *theorize* that you believe there is a party on Friday night and desire to go to the party, and that you do not have a competing desire not to attend the party. Thus, I predict that you will go to the party. Suppose that my reasoning process just so happens to resemble your reasoning process. We both weigh various considerations, decide how to maximize preferences, etc. Despite the fact that I am *theorizing* about your behavior, my cognitive system merely *abstractly* replicates your decision-making process. My reasoning process only *accidentally* resembled your reasoning process.[4] You may have employed a very different sort of reasoning to get to the conclusion that you will go to the party. The success of my prediction does not *depend* on a fine-grained similarity between my cognitive process and yours. Thus, the TT entails abstract replications.

Given that abstract replications are so easy to come by, and that the TT entails them, the ST needs a more restrictive account of simulation. The ST needs to be restricted to concrete replications, which are not ubiquitous and not entailed by the TT. This means that simulation theorists must show that their examples of

[4] It would count as abstract replication even if the resemblance were non-accidental. I emphasize accidental resemblance in the main text to highlight the fact that the success of an abstract replication does not depend on any fine-grained similarities between the processes. This is further evidence that abstract replication is too liberal for ST.

simulational mindreading are cases in which the mindreader's mental process replicates the actual mental process the target goes through, and that the non-accidental success of the simulation *depends on* replicating that mental process. This is not an overly burdensome requirement, but it will limit what counts as simulational mindreading (Fisher, 2006).

Goldman's second criterion for a process P to simulate P´ is that in its duplication of P´, P fulfills one of the purposes or functions of P. How can we tell when a process has the function of (concretely) replicating another process? Simulation theorists do not provide many details, but they give examples of what they take to be simulational processes. These examples include using one's own inference mechanisms to determine what a target believes (Nichols & Stich, 2003, p. 135), using one's own inference mechanism to predict the kind of inferences others will make (Stich & Nichols, 1995); using one's own grammaticality system to determine whether a target will judge a particular sentence to be grammatical (Harris, 1995); using our knowledge of a scientific theory to predict what other experts in the field believe (Heal, 1995); using our arithmetical skills to figure out how a target will answer an arithmetic question (Heal, 2003).

The problem with these examples is that they do not discriminate cases in which a process *happens* to replicate another process from cases in which replication is the *function* of the process. Consider the following case, which shares the same form as the other examples but is not a case of simulational mindreading (Ramsey, 2010). In watching a baseball game, I predict that the pitcher will throw to first base. I see that the runner has strayed too far off base, I see that the pitcher sees this, and I infer that since the pitcher sees that the runner has strayed too far off base that he believes this. This belief, plus the attributed belief that runners who stray too far off base are easy to throw out, and the attributed desire to throw out runners, lead me to predict that the pitcher will throw to first base. I theorize about what the pitcher will do and why. When I see what the pitcher sees, I reason as the pitcher reasons. My perceptual system generates a belief about the status of the runner, as does the pitcher's perceptual system. But this does not mean that I am simulating the pitcher's visual system and decision-making processes. In this case I am using my visual system in the way that it is normally used; to ascertain various facts about the world that I can use in my reasoning. I employ these facts in reasoning about where the pitcher ought to throw. The fact that the states of my perceptual system and reasoning process about where the ball needs to be thrown resemble the processes taking place in the pitcher's mind does not entail that I am simulating the pitcher. I am simply using my cognitive capacities as I normally do, i.e., as a fact finder, and on the basis of the facts I know about the world and the pitcher, I predict that the pitcher will throw to first base (Ramsey, 2010).

We must distinguish when our cognitive mechanisms are being used as they normally are – as *fact finders* that merely happen to replicate the cognitive processes in a target – from when they are being used to simulate a target's cognitive processes. In other words, we need a way to determine when a process has the *function* of concretely replicating another process. In order to determine this, we need to look at two features of the cognitive process in question: how it operates and its content (Nichols & Stich, 2003, pp. 132-135; Ramsey, 2010).

Genuine mental simulation requires that the cognitive process be taken "offline" and provided with non-standard inputs.[5]

The first criterion for a process P to have the function of replicating P` is that P is taken "offline" in its simulating role. When I am genuinely simulating a target's cognitive processes, I use my own cognitive mechanisms but the process does not result in an action, decision, etc., as it does when I am using my cognitive mechanisms "online." Rather, it is merely an imagined action, decision, etc. Importantly, this criterion is neutral with respect to whether P in its offline role is conscious and voluntary or unconscious and involuntary.

The second criterion is that the content (i.e., the input and output) of my simulational cognitive process is (or aims to be) the content of the target's cognitive process. In other words, inputs and outputs of the simulational process are non-standard; they are imaginative representational surrogates for the target's mental states. This criterion is neutral with respect to the kind of representational surrogates involved. These may be propositional attitudes or non-propositional mental states like emotions or images.

These two criteria for figuring out whether P has the function of simulating P` – offline operation and representational surrogates – allow for processes that are conscious, voluntary, and involve propositional attitudes *and* processes that are unconscious, non-voluntary, and involve non-propositional attitudes. This means

[5] Though these criteria are standard for ST, Goldman has argued that they are required only for high-level simulation (Goldman, 2006b, p. 131). I consider this argument in section 4 below.

that these requirements for simulation apply both to *low-level* and *high-level* mindreading.

Putting this all together, we have the following definition of simulation:

1. P concretely resembles P′, which requires that:
    a. P and P′ involve the same type of system;
    b. P and P′ exhibit similar fine-grained details.
2. In resembling P′, P fulfills one of its functions, which requires that:
    a. The success of the simulation depends on P concretely resembling P′;
    b. P is taken "offline" in its simulating role;
    c. The input and output of P are representational surrogates for the input and output of P′.

Consider again some of the examples discussed so far. On this definition, the way in which I determined that you would attend the party Friday night – theorizing about what you would do in a way that just so happened to resemble your own reasoning process – does not count as a simulation because it does not meet 2a, 2b, or 2c. My prediction that the pitcher will throw to first base is not the result of a simulation because it does not meet 2a, 2b, or 2c, either. A legitimate case of simulation is one where I use my cognitive system to concretely replicate your mental process, i.e., I adopt representational surrogates for your mental states and run them through my own offline cognitive mechanisms, and the success of this process depends on replicating the mental process that you actually go through.

We ought to adopt this more precise conception of simulation. Though it is more demanding than Goldman's generic conception discussed above, the problem with the generic conception is that it is far too lax. This more precise account of simulation rules out those illegitimate examples of simulation and captures what is truly distinctive about legitimate cases of simulation. Moreover, it offers a unified conception of simulation that is appropriate both for high-level and low-level simulation.

3. High-Level Simulation

Now that we have clarified the concept of simulation, we can discuss high-level and low-level simulation. Before philosophers and cognitive scientists conceived of the concept of low-level mental simulation, all simulational mindreading was considered high-level, i.e., consciously accessible and voluntarily mediated. In this section, I describe exactly how high-level simulation mindreading is meant to work and identify some theoretical worries about it.

Recall the two kinds of simulational mindreading. First the subject retrodictively simulates the target in order to figure out what mental states could have caused that behavior, and then the subject projectively simulates in order to figure what the target will do next. Suppose I have a guest lecturer in my class one day and a student comes in late. She slowly opens the door, looks to the front of the room, and leaves. According to the ST, first I imagine myself in the student's situation. I imagine that I am coming in late to class, I see a stranger at the front of the room, and I walk out. I imagine what I would think in this situation. Perhaps I would think that in my rush to get to class, I walked into the wrong classroom. Or maybe I would think a guest lecturer is not worth my time, so I leave. Or perhaps I know and dislike the guest lecturer, so I prefer not listen to him. I imaginatively adopt these mental states to see if they would explain the observed behavior. I determine that if I were in that situation, I would think that I walked into the wrong class. Thus, I attribute to the student the desire to attend my class and the false belief that she has walked into the wrong classroom.

For the projective simulation, the subject takes the results of the retrodictive simulation and imagines what she would do next if she were in the target's situation. In our example, I imaginatively adopt the desire to attend class and the belief that I have walked into the wrong classroom, and I run this information through my own decision-making mechanisms. I imaginatively decide to check the classroom location on the syllabus, which as a good student I always carry with me. I imagine that having done this, I would realize that I did not walk into the wrong classroom. I imaginatively decide to peek into the classroom again to see if I recognize the other students in the room. I imagine that doing this would result in the realization that all of my classmates are in the room, and that the person at the front of the room is just a guest lecturer, and as a result I would come back into the classroom. Thus, through this simulational process, I generate a prediction about what the student will do. I predict that she will check her syllabus, realize that she did not make a mistake, and then come back into the classroom.

This example highlights several facts about high-level simulational mindreading. First, high-level simulation involves imagination in the conventional sense. In both the retrodictive and projective simulation, I consciously and deliberatively imagine myself in the target's circumstances. Second, high-level simulation easily can be co-opted to explain our engagement with fiction (Currie & Ravenscroft, 2002; Doggett & Egan, 2012). I simply put myself in the fictional character's position and imagine what I would think, feel, and do in that situation. I imaginatively adopt the beliefs and desires of the character and in some circumstances even have affective responses to the fictional events. When Othello approaches Desdemona in a murderous rage, I feel something like fear horror at this impending fictional murder.[6] Third, this account of simulational mindreading could be co-opted to explain how some thought experiments work (Williamson, 2015). In some thought experiments, you imagine being in a particular situation and then you entertain a question about what you would think or do in that situation. In this way, potentially you could get knowledge through simulational imagination.[7]

One worry about the ST is about whether it is computationally tractable. For ST, explanation requires identifying what the possible relevant mental states of the target could have been in order to cause the observed behavior, running the imagined mental states through one's own (offline) decision-making mechanism

[6] The characterization of these affective responses is a matter of much dispute. Some argue that the affective responses simply are emotions (Carroll, 1990; Gendler, 2008; Weinberg & Meskin, 2006), where as others argue that they are best characterized as imaginative emotions (Goldman, 2006a; Velleman, 2000; Walton, 1990). Simulation theorists tend to fall in the latter category, but we need not take a stance on that issue here.

[7] See Spaulding (2015) for a discussion of how exactly we can and do get knowledge through imagination.

and, once plausible candidate mental states are identified, attributing the computed imagined decision to the target as an explanation of past behavior and prediction for future behavior. For social interactions that involve a range of people and behaviors, one would first need to identify the possible relevant mental states of all the participants and run the simulations simultaneously to come up with a coherent explanation of the participants' behaviors and then use this explanation to imaginatively simulate the decisions of the participants. It is an open question whether it is possible to run multiple simultaneous simulations through one's own decision-making mechanism. However the more serious problem is that the targets' behavior and mental states will be interdependent. ST's simultaneous, interdependent simulations, if they are even possible, would be computationally intractable (Bermúdez, 2003a, pp. 31-34).

Another theoretical limitation of the ST is what is known as the "threat of collapse" (Davies & Stone, 1995b). To retrodictively simulate my student, I observe her walking out of class upon seeing the guest speaker, imagine myself her situation, generate hypothetical beliefs and desires that would explain why I would walk out of class if I were in that situation, and then attribute those mental states to my student. The difficulty here is that there are indefinitely many mental state combinations that would explain the observed behavior. If we were to try to figure out, with simulation resources only, what our mental states could have been to cause us to behave like the target, our retrodictive simulation would have no way to decide between radically different belief-desire combinations that would explain the behavior. Moreover, there would be no stopping point for the retrodictive simulation. The simulation itself provides no way to determine when we have landed on a good-enough explanation of the observed behavior and can stop simulating.

This problem is called the threat of collapse because, upon inspection of imagination-based simulation, it is evident that the ST needs the theoretical knowledge posited by the TT. Thus, it is argued, perhaps hyperbolically, that the ST simply collapses into the TT. Retrodictive simulation reveals some of the possible mental states that a target may have, but it cannot, all by itself, provide knowledge of other minds. Theoretical information is required to move from identifying possible mental states to knowing a target's mental states. Most simulation theorists these days recognize this problem and as a result endorse hybrid ST/TT models (Davies & Stone, 1995b; Goldman, 2006b; Heal, 1998; Nichols & Stich, 2003). The hybrid models handle the threat of collapse, and they may also alleviate the computational worries about running multiple, interdependent simulations.

4. Low-Level Simulation

Like high-level simulation, low-level simulational mindreading involves both retrodictive and projective simulation. One central difference between high-level and low-level simulation is that one has neither conscious access to nor control over the process or product of low-level simulational mindreading. Low-level simulation involves an unconventional conception of imagination. In the context of low-level simulation, imagination operates unconsciously and automatically.

The ordinary, folk conception of imagination is that it is conscious and potentially under the agent's voluntary control. However, several theorists have argued that imagination may be non-conscious and non-voluntary (Church, 2008; Goldman, 2006b; Nanay, 2013; Van Leeuwen, 2014; Walton, 1990). Kendall Walton, for example, posits non-occurrent imaginings that do not occupy the subject's attention. Bence Nanay's conception of pragmatic mental imagery involves spatially rich representations that may or may not involve conscious visualization. Jennifer Church argues that several ordinary cases of imagining are most plausibly explained by unconscious imagining. She considers several objections to the idea that imagination can be non-conscious and develops her own account of unconscious imagining. In each of these cases, imagination involves imagery, but this imagery may not be conscious and it need not be under the voluntary control of the subject. Although the notion of unconscious imagining is unconventional, it is not unprecedented. Several prominent figures in the philosophy of imagination argue for it. Thus, it is appropriate to hold that low-level simulation involves imagination.

I argued above that high-level and low-level simulation have the same basic requirements. It is worth pausing to consider an objection to my claim that the above requirements apply to high-level *and* low-level simulation. Goldman argues that 2b (offline operation) and 2c (representational surrogates) are required only for *high-level* simulational mindreading. These two criteria require using pretense or imagination, which Goldman argues are high-level activities that are potentially and intermittently under intentional guidance or control. Low-level simulational mindreading is supposed to be fully automatic, though (Goldman, 2006b, pp. 131-132).

The definition of simulation I offer can accommodate Goldman's claims about low-level mindreading. Imagination is required to produce representational surrogates for the low-level simulation, but above I cite several prominent theorists who allow for unconscious, involuntary imagination. Even by Goldman's own lights, imagination need not be a conscious, voluntary process. In characterizing *E-imagination* (his prototype for high-level simulational mindreading), Goldman cautions against adopting the naïve conception of imagination, which is restricted to the sphere of consciousness and the control of the will. "E-imagination is introduced here as a psychological construct, the referents of which can either be conscious or covert, voluntary or automatic, and these properties can hold for both the generating process and the products so generated" (Goldman, 2006b, p. 151). We can say that low-level simulation involves psychological surrogates (non-consciously and automatically produced by imagination) that are run through one's offline cognitive mechanisms. Thus, these criteria for simulation can be applied to low-level simulation, as well.

Moreover, rejecting the stricter criteria comes at a significant cost. If we adopt Goldman's generic definition of simulation for high-level and low-level simulation, then simulation just is process resemblance. However, process resemblance does not discriminate simulational processes from non-simulational processes. Predicting that you will go to the party or that the pitcher will throw to first base both involve process resemblance, but these are cases of theorizing, not simulating. Adopting the generic conception for high-level and low-level simulation would obscure the distinction between simulation and theory.

Alternatively, if one adopts the stricter criteria for high-level simulation but the generic definition for low-level simulation, then low-level and high-level simulation would end up looking quite different. High-level simulation would involve concrete resemblance, offline processing, and representational surrogates, whereas low-level simulation would involve none of these features. Low-level simulation simply would be process resemblance. The criteria for low-level simulation would be significantly weaker and different than the criteria for high-level simulation. Adopting different criteria for low-level and high-level simulation threatens the idea that there is a unified conception of simulation. "Low-level simulation" would be a different sort of thing altogether, and evidence of "low-level simulation" would not be evidence for the ST in general. Fragmenting simulation in this way would seriously damage the ST. Thus, there are several good reasons to adopt the stricter criteria for high-level *and* low-level simulation.

By far the most discussed example of low-level simulational mindreading is mirror neurons. Mirror neurons are neurons that fire, or activate, when a subject acts, emotes or experiences a certain sensation, and also when a subject observes a target acting, emoting or experiencing a certain sensation. For example, a host of neurons in the premotor cortex and posterior parietal cortex fires when I grasp an object, and this same host of neurons fires when I observe another person grasping an object (Rizzolatti & Craighero, 2004). There are similar mirror neuron systems for experiencing and observing certain emotions. When I experience disgust and when I observe another person experiencing disgust the same collection of neurons in the insula fires (Calder, Keane, Manes, Antoun, & Young, 2000; Wicker et al., 2003). Similar findings hold for the experience and observation of fear (Adolphs, Tranel, Damasio, & Damasio, 1994), anger (Lawrence, Calder, McGowan, & Grasby, 2002), pain (Singer et al., 2004) and touch (Keysers & Perrett, 2004). In each of these cases, groups of neurons are endogenously activated when the subject acts, emotes, or feels a certain way, and these same groups of neurons are exogenously activated (at an attenuated level) when the subject observes another acting, emoting, or feeling in those same ways.

Some theorists propose that mirror neurons are the basis for our ability to understand and interact successfully with other people (Gallese, 2007; Gallese, Keysers, & Rizzolatti, 2004; Hurley, 2005; Keysers & Gazzola, 2009; Oberman & Ramachandran, 2009). The argument goes like this. How is it that we understand what other people are doing, why they are doing it, what they are going to do next? Mirror neuron studies demonstrate that parts of our brains fire in the same ways when we observe an action, emotion, or sensation and when we act, emote, or sense in the same way. The suggestion is that we understand what another person is doing, feeling, and experiencing because when we observe the other person parts of our brains are activating as if *we* were doing what the other person is doing. Our brain activity mirrors – that is, simulates – the other person's brain activity such that it is *as if* we are acting, feeling, or experiencing how the target is acting, feeling, or experiencing.

Controversy surrounds many aspects of mirror neurons (Hickok, 2009). In fact, it is not even clear whether the empirical facts support the claim that mirror neurons are simulational in the relevant respect (Herschbach, 2012; Spaulding, 2012, 2013). In this context, however, we need not concern ourselves so much with these controversies. Mirror neurons are a common example of how low-level simulation works. Mirror neurons are supposed to be an example of a neural mechanism that has the function of concretely replicating a target's experience. We do not need to worry about whether this particular neural mechanism actually meets all the criteria in order to use it as an illustrative example.

Low-level simulation faces a version of the threat of collapse (Spaulding, 2012, 2015). The threat of collapse for high-level simulation is that indefinitely many propositional attitude combinations are compatible with the observed behavior, and simulation itself provides no way to determine when we have landed on a plausible, good-enough explanation of the observed behavior. The threat of collapse for high-level simulation concerns propositional attitudes. Low-level simulation involves non-propositional attitudes, such as basic intentions and emotions. I will use mirror neurons to illustrate the threat of collapse for low-level simulation. Again, though there are genuine concerns about whether the empirical data supports the claim that mirror neurons are simulational in the relevant respect, let us set aside those concerns for now.

The threat of collapse for low-level simulation is that an observed behavior or facial expression is compatible with a number of different basic intentions or emotions. A blush may indicate embarrassment, happiness, anger, or even just a hot flash. The same applies even more clearly for basic intentions. A given behavioral movement may indicate an intention to eat, give, tease, throw, play with, put away, etc. We need more information than the simulation heuristic provides in order to be justified in attributing to a target a particular intention or emotion.

The observed behavior is compatible with a number of mental states, and the low-level simulation itself provides no way to determine the plausibility of the candidate emotions or intentions. That is, there is no sufficing heuristic or stopping point built into the operation of the low-level simulation. We need other information to discriminate among the intentions that could cause the behavior, or the emotions that could cause the blush. This may be information about the target's recent history, her personality, how certain situations make her feel, folk psychological platitudes about how behaviors relate to mental states, etc.

One way to mediate the threat of collapse for high-level simulation is to endorse of a hybrid ST/TT model. The TT provides the theoretical information required to move from identifying possible mental states to knowing a target's mental states. There is an analogous move for low-level simulation. One could hold that low-level simulational mechanisms receive input from and send output to non-simulational mechanisms. In fact, functional anatomical evidence supports the idea that simulational and non-simulational mechanisms are integrated in this way.

During action observation, mirror neurons receive input from the superior temporal sulcus (STS), an area independently associated with perception of social stimuli, mentalizing, and action understanding (Rizzolatti & Craighero, 2004, p. 172). Mirror neuron regions send output to the primary and secondary somatosensory cortices and the middle temporal lobe (Rizzolatti & Sinigaglia, 2010, p. 265). The STS, mirror neuron areas, and somatosensory cortices are informationally integrated. These neural regions, in addition to the aFMC, posterior STS, and temporal parietal junction (TPJ), are involved perception of social stimuli, action understanding, and mentalizing (Brass, Schmitt, Spengler, & Gergely, 2007).

These data show that there are several neural regions informationally integrated with mirror neurons that are involved with the same functions that mirror neurons are alleged to be involved in. Importantly, these other regions do not operate by neural reuse. Thus, one empirically and theoretically plausible way to address the threat of collapse for low-level simulation is to conceive of the low-level simulational mechanism as informationally integrated with other non-simulational mechanisms. These non-simulational mechanisms supplement the simulational mechanism in the same way that theoretical knowledge supplements high-level simulation.

5. Conclusion

There has been much philosophical and empirical work on the ST in the last decade or so, and as a result it is a flourishing theory. My discussion in this chapter attempts to clarify some aspects of the ST, specifically, the concept of simulation, high-level simulation, and low-level simulation. I argued that we have good reasons to adopt a more precise conception simulation for high-level and low-level simulation. The stricter criteria rule out some illegitimate examples of simulation, and they also capture what is truly distinctive about legitimate cases of simulation. I hope this more precise conception of simulation will contribute to the flourishing of the ST.

References:

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature, 372*(6507), 669-672.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17*(24), 2117-2121.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience, 3*(11), 1077-1078.

Carroll, N. (1990). *The Philosophy of Horror*. New York: Routledge.

Carruthers, P., & Smith, P. K. (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Church, J. (2008). The Hidden Image: A Defense of Unconscious Imagining and Its Importance. *American Imago, 65*(3), 379-404.

Currie, G. (2010). Tragedy. *Analysis, 70*(4), 632-638.

Currie, G., & Ravenscroft, I. (2002). *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Oxford University Press.

Davies, M., & Stone, T. (1995a). *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell.

Davies, M., & Stone, T. (1995b). *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.

Doggett, T., & Egan, A. (2012). How We Feel About Terrible, Non‐existent Mafiosi. *Philosophy and Phenomenological Research, 84*(2), 277-306.

Fisher, J. C. (2006). Does simulation theory really involve simulation? *Philosophical Psychology, 19*(4), 417-432.

Gallese, V. (2007). Before and below theory of mind: embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society, 362*(1480), 659-669.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*(9), 396-403.

Gendler, T. S. (2008). Alief and belief. *Journal of Philosophy, 105*(10), 634-663.

Goldman, A. I. (2006a). Imagination and Simulation in Audience Responses to Fiction. In S. Nichols (Ed.), *The Architecture of the Imagination* (pp. 41-56). Oxford: Oxford University Press.

Goldman, A. I. (2006b). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*: Oxford University Press, USA.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language, 1*, 158-171.

Harris, P. (1995). From simulation to folk psychology: The case for development. In M. Davies & T. Stone (Eds.), *Folk Psychology* (pp. 207-231). Oxford: Blackwell.

Heal, J. (1995). How to think about thinking. In M. Stone & M. Davies (Eds.), *Mental Simulation: Evaluations and Applications*: Blackwell.

Heal, J. (1998). Co-cognition and off-line simulation: Two ways of understanding the simulation approach. *Mind & Language, 13*(4), 477-498.

Heal, J. (2003). *Mind, Reason and Imagination*. Cambridge: Cambridge University Press.

Herschbach, M. (2012). Mirroring versus simulation: on the representational function of simulation. *Synthese*, 1-31. doi: 10.1007/s11229-011-9969-6

Hickok, G. (2009). Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. *Journal of Cognitive Neuroscience, 21*(7), 1229-1243. doi: 10.1162/jocn.2009.21189

Hurley, S. (2005). The shared circuits hypothesis: A unified functional architecture for control, imitation, and simulation. In S. Hurley & N. Chater (Eds.), *Perspectives on Imitation: From Neuroscience to Social Science* (pp. 177): MIT Press.

Keysers, C., & Gazzola, V. (2009). Unifying social cognition. In J. A. Pineda (Ed.), *Mirror Neuron Systems* (pp. 1-35). New York: Humana.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences, 8*(11), 501-507.

Lawrence, A. D., Calder, A. J., McGowan, S. W., & Grasby, P. M. (2002). Selective disruption of the recognition of facial expressions of anger. *Neuroreport, 13*(6), 881-884.

Nanay, B. (2013). *Between perception and action*: Oxford University Press.

Nichols, S., & Stich, S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

Oberman, L. M., & Ramachandran, V. S. (2009). Reflections on the mirror neuron system: Their evolutionary functions beyond motor representation. In J. A. Pineda (Ed.), *Mirror Neuron Systems* (pp. 39-62). New York: Humana.

Ramsey, W. M. (2010). How not to build a hybrid: Simulation vs. fact-finding. *Philosophical Psychology, 23*(6), 775-795.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*(1), 169-192.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience, 11*(4), 264-274.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162.

Spaulding, S. (2012). Mirror neurons are not evidence for the Simulation Theory. *Synthese, 189*(3), 515-534.

Spaulding, S. (2013). Mirror neurons and social cognition. *Mind & Language, 28*(2), 233-257.

Spaulding, S. (2015). Imagination Through Knowledge. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*: Oxford University Press.

Stich, S., & Nichols, S. (1995). Second thoughts on simulation. In M. Davies & M. Stone (Eds.), *Mental Simulation: Evaluations and Applications*: Oxford.

Van Leeuwen, N. (2014). The Meanings of "Imagine" Part II: Attitude and Action. *Philosophy Compass, 9*(11), 791-802. doi: 10.1111/phc3.12141

Velleman, J. D. (2000). On the aim of belief *The Possibility of Practical Reason* (pp. 244–281). Oxford: Oxford University Press.

Walton, K. L. (1990). *Mimesis As Make-Believe: On the Foundations of the Representational Arts*: Harvard University Press.

Weinberg, J. M., & Meskin, A. (2006). Puzzling Over the Imagination: Philosophical Problems, Architectural Solutions *The Architecture of the Imagination*. Oxford: Oxford University Press.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664.

Williamson, T. (2015). Knowing by imagining. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*. Oxford: Oxford University Press.